

第6章 補論

音声認識を用いた情報保障システム運用の課題

坂本徳仁^(†)

本補論では、前章で説明した音声認識を用いた情報保障システムに関する内容について、シンポジウム「聴覚障害者の情報保障を考える」報告後に得られた研究成果と現時点での課題について簡単に説明する。

最初に、原文復唱方式と要約復唱方式の間の認識精度および誤認識等の違いについて実験を行なったところ、やはり原文復唱方式の方が要約復唱方式よりも認識精度が5～10%程度高いという結果が得られた(坂本・櫻井・鹿島2010)。この結果、原文復唱方式は「認識率は高いが、話し言葉をそのまま字幕化するため、提示文章が分かりにくい」という性質を有し、要約復唱方式は「提示文章は分かりやすいが、認識率は低い」という正反対の性質をもつことが明らかになっている。残念ながら、現時点で、認識精度を上げつつ、分かりやすい提示文章を作成する方法について、筆者らの研究グループは良い解決策を持ち合わせていない。今後は要約筆記者の要約技術を参考にしながら、要約復唱入力方式の認識精度を高めるような方法を探求していく必要がある。

続いて、筆者らの研究グループは、筑波技術大学の研究グループが中心となって開発した音声認識を用いた情報保障システム(「音声認識によるリアルタイム字幕作成システム構築マニュアル」編集グループ[編][2009]を参照のこと)を立命館大学に移植し、試験的に運用した。その結果、音声入力の認識精度も字幕提示にかかるタイムラグも先行研究と同水準の成績(認識精度8～9割台、タイムラグ10秒前後)を得られるようになったが、システムを運用するに当たって非常に大きな二つの障害があったため、現状での実用化は困難であると判断している。ここで、実用化を困難にしている二つの障害とは、①復唱者の養成・

表 1 各方式の性能比較

(1) 音声入力方法による分類

	話者入力方式	復唱入力方式
認識精度	低い (5～7割)	高い (8割台、訓練すれば9割台)
人件費	安い (校正者2名)	高い (校正者2名+復唱者2名)
人材育成	容易 (校正者のみ)	困難 (復唱者の育成が困難)

(2) 復唱方式による分類

	原文復唱方式	要約復唱方式
認識精度	高い (7～8割台)	低い (6～8割台)
提示文章	わかりにくい	わかりやすい
人材育成	相対的に容易 (復唱作業のみ)	困難 (復唱+要約作業が困難)

(3) 復唱者の位置関係による分類

	近接方式	遠隔方式
運用システム	簡素	やや複雑
ノイズ状況	悪い	良い
会話の進行	妨げになる	無関係

(4) 校正方式による分類

	音声遅延方式	録音方式
誤字修正精度	やや低い (95%以上)	高い (ほぼ100%)
字幕化までのタイムラグ	小さい (10秒程度)	大きい (20秒以上)

確保の問題と、②人件費の問題である。本書の第6章2節の議論でも触れたが、復唱作業は一般に難しく、上手にできる人はほとんどいない。筆者の感触では、話者がどんなスピードであっても復唱を上手にできる人は10人中1人もいればよい方である。しかも、復唱作業は要約筆記や手話通訳と同じく15～20分間隔で交代する必要があるため、復唱入力方式を採用する場合には少なくとも2名の復唱者が必要とされている。したがって、大学の講義や企業内での会議・研修といった場での情報保障において、復唱入力方式を採用した音声認識字幕化システムを安定的に運用できるだけの十分な数の復唱者を確保することは相当程度の困難を伴う。さらに、現状では、システムの運用に最低限必要な人数

が復唱者2名と校正者2名の合計4名となっているため、通常の手話通訳や要約筆記を用いた情報保障システムと比べて2倍以上の人件費がかかってしまう。現状では、これら二つの問題を克服することができなかったため、残念ながら、音声認識を用いた情報保障システムを実用するにはいたっていない。

以上、前章までの議論と本補論で報告した結果を合わせると、音声認識を用いた情報保障システムの諸問題は表1のようにまとめることができる。

最後に、音声認識を用いた情報保障システムはパソコン要約筆記よりも多くの文字を提示することができるため、音声情報の欠落が少なく済むという利点をもっている。したがって、音声認識の技術が今後進歩していくのであれば、音声認識を用いた情報保障システムは情報保障の手段として非常に有望なものとなろう。音声認識の専門家である東京工業大学古井貞熙教授によれば、音声認識の技術はまだ道半ばのもので、完成に至るまでの道のりはまだ半分近くも残っているようである（古井 2009）。話者から直接音声を入力しても十分な認識精度が得られ、結果として校正者の負担も少なくなれば、人件費がほとんどかからない形での音声認識による字幕化の運用も可能となる。その意味において、音声認識を用いた情報保障システムが安定的かつ費用節約的に運用できるか否かは、現時点では音声認識の技術水準に全面的に依存していると言っても過言ではなく、人材育成やシステム開発といったソフト面での問題よりも音声認識技術という意味でのハード面の問題の方が圧倒的に重要であるように思われる。

【参考文献】

「音声認識によるリアルタイム字幕作成システム構築マニュアル」編集グループ [編] (2009) 『音声認識によるリアルタイム字幕作成システム構築マニュアル』, 日本聴覚障害学生高等教育支援ネットワーク.

坂本徳仁, 櫻井悟史, 鹿島萌子 (2010) 「音声認識字幕化システムにおける要約・原文復唱入力方式の比較分析」, 障害学会第7回大会ポスター報告, 東京.

古井貞熙 (2009) 『人と対話するコンピュータを創っています——音声認識の最前線』, 角川学芸出版.